WPI

Major Qualifying Project

MAPPING DNA REPLICATION IN THE YEAST SPECIES *SACCHAROMYCES CEREVISIAE*

Anna Wortman

Date: 4/27/2017

MAPPING DNA REPLICATION IN THE YEAST SPECIES *SACCHAROMYCES CEREVISIAE*

A MAJOR QUALIFYING PROJECT SUBMITTED TO THE FACULTY OF WORCESTER POLYTECHNIC INSTITUTE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE

Submitted By:

Anna Wortman

Submitted to:

Advisor: Professor Reeta Rao

Date: 4/27/2017

ABSTRACT

DNA replication is an essential function that allows all living organisms to maintain life. This function is coordinated with other aspects of genome metabolism such as DNA repair, chromatin structure, and gene expression. Timing is important in this coordination- the timing of replication origin firings drives replication timing. One way timing is regulated is by loading of the replicative helicase complex, MCM. An origin is more likely to fire earlier when more MCM, a hexamer of six polypeptides that aids in the formation and elongation of the replication fork, is loaded by the Origin Recognition Complex (ORC).

The goal of this MQP is to examine the number of MCM complexes loaded on single replication origins before S phase, the synthesis phase of DNA replication, is initiated. In order to identify the specific number of MCM complexes that are loaded on any given origin, a *Saccharomyces cerevisiae* strain was built that contains a TALO8 plasmid with a single origin, a LacI-SNAP cassette to tether TALO8, and a MCM4-fluorescence cassette to microscopically track MCM. This plasmid will be captured in a flow cell and microscopically studied in order to count the number of MCMs loaded on single plasmids containing known origins.

ACKNOWLEDGEMENT

I would like to thank Dr. Nick Rhind for giving me the opportunities to volunteer in his lab and inviting me to do this research project. His support, guidance, and advice have helped me immensely throughout this project. I would also like to thank Livio Dukaj for assisting me throughout the laboratory procedures and taking the time to answer all of my questions, I couldn't have completed this project without him. I'd like to extend thanks to my advisor, Professor Reeta Rao, for guiding me through this process. Finally, I would like to express my gratitude for the support from my family and friends. Without all of their help, I would not have been able to come so far in my academic journey.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

INTRODUCTION

Replication timing

DNA replication in eukaryotic cells is limited to a specific window of time, known as S phase. To advance successfully through this phase, the entire genome must be copied correctly, exactly once, and within a timely manner; otherwise, errors in replication could lead to genome instability and cell death (Bell & Kaguni, 2013) (Rhind & Gilbert, DNA replication timing, 2013). To this end, DNA replication has evolved to be a precisely coordinated process, dependent on an ordered series of steps in order to produce the factors necessary for all phases of the cell cycle.

The process of eukaryotic replication begins from several locations, or origins of replication, on each chromosome. The timing of replication origin firing determines replication timing (Goren & Cedar, 2003). Occurring during the G1 phase of the cell cycle, replication origin loading begins when the Origin Recognition Complex, or ORC, binds to the genome's origins and loads the eukaryotic replicative helicase complex, minichromosome maintenance complex (MCM), around the DNA (Sakakibara, Kelman, & Kelman, 2009) (Yeeles, Deegan, Janska, Early, & Diffley, 2015).

MCM protein is made up of about 650 amino acids structurally divided into three parts: an N-terminal, a catalytic region, and a C-terminal helix-turn-helix domain (Sakakibara, Kelman, & Kelman, 2009). A heterohexameric MCM2-7 helicase complex forms at replication origins to unwind double stranded DNA and power fork progression (Figure 1) (Brewster, et al., 2008).

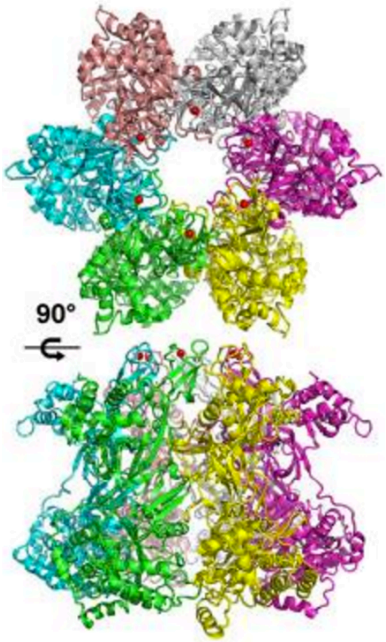

Figure 1. Ribbon diagram showing the top and side views of a hexamer model of MCM. Reprinted from Brewster et.al. (2008)

Once the MCM complex is activated, replication is induced and replication timing for the genome is determined. Within this phase, some portions of the genome replicate early, and others late, creating a characteristic pattern of replication timing. Recent models propose stochastic regulation of origin firing wherein firing time of an origin within a population is equivalent to the probability of that origin firing at the single-cell level (Yang, Rhind, & Bechhoefer, 2010) (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015). An origin that fires with high probability is more likely to fire early in S phase while an origin that fires with lower probability is unlikely to fire early in S phase, will have a later average firing time and will be passively replicated (Yang, Rhind, & Bechhoefer, 2010) (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015).

As cells differentiate, origin firing patterns change and correspond to patterns of transcriptional regulation and chromosome structure, implying a relationship between replication timing and processes of genome metabolism like gene expression and genome evolution (Goren & Cedar, 2003) (Rhind & Gilbert, DNA replication timing, 2013). However, there are uncertainties in the regulation of origin firing, leading to two different hypotheses for

the stochastic model (Rhind, Yang, & Bechhoefer, Reconciling stochastic origin firing with defined replication timing, 2010).

Stochastic Model of Replication

The stochastic model of replication suggests that the firing time of an individual origin of replication in a population is heterogeneous, this model assumes that the firing of one origin does not affect another and that the process is independent (Bechhoefer & Rhind, 2012). Studies have demonstrated heterogeneous patterns of origin firing (Patel, Arcangioli, Baker, Bensimon, & Rhind, 2006). From them, two central conclusions have been drawn- first, the firing of eukaryotic replication origin firing is inefficient and second, stochastic. The inefficiency of eukaryotic origins is well documented (Kalejta & Hamlin, 1996) (Czajkowsky, Liu, Hamlin, & Shao, 2008). Some yeast origins are as high as 90% efficient and others less than 10% (Hiraga, et al., 2014). This inefficiency suggests some origins are passively replicated by a fork from another nearby origin; thus, the longer an origin goes without being passively replicated, the better chance it has of firing independently.

Stochastic Model Evidence

Experimentally, the stochastic model has been addressed in a large population of cells based on mathematical analysis of replication kinetics. This model, based on the budding yeast *Saccharomyces cerevisiae*, outlines a multiple-MCM system in which replication timing is dependent on the number of MCM complexes loaded on origins of replication (Yang, Rhind, & Bechhoefer, 2010) (Rhind, Yang, & Bechhoefer, Reconciling stochastic origin firing with defined replication timing, 2010). Following this model, MCM complexes are stochastic and activate with similar probabilities across the genome. Origins with more MCM complexes loaded are, on average, more likely to fire early in S phase (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015).

The multiple MCM model was first used to show that early-firing origins have more MCM complexes loaded than do later-firing origins. Using ChIP-seq in G1 arrested cells, the genome-wide distribution of MCM complexes was examined. The signal was concentrated on known origins, with increased levels on those that fire earlier: ARS1012, ARS1014, ARS1018,

and ARS1019 (Fig. 2A). This data showed that there is a strong correlation between the ChIP-seq signal and the origin timing parameter, n, across the genome (Fig. 2B) (Yang, Rhind, & Bechhoefer, 2010). N is a direct estimate of origin timing, describing the firing-time distribution of origins and not origin replication, which is influenced by passive replication (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015). The origins whose signal falls above the diagonal represent more MCM complexes loaded than the mathematical model predicted. These include telomeric origins, which are known to fire late in a hetero- chromatin-dependent manner and those delayed in firing by Rpd3 HDAC, a histone deacetylase that removes lysine residues on the N-terminal part of the core histones and prevents transcription (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015) (UniProtKB - P32561 (RPD3_YEAST)).

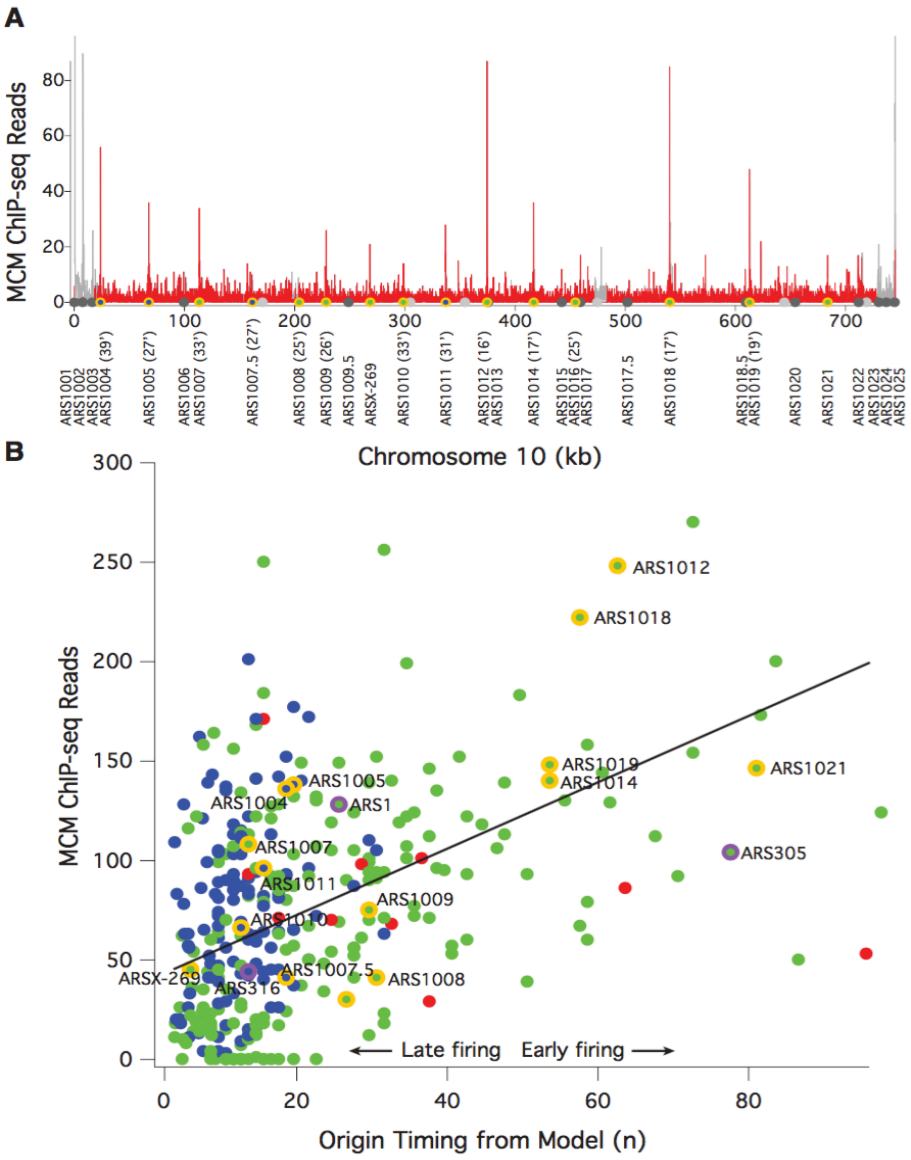

Figure 2. MCM ChIP-seq data. Reprinted from Das et. al. (2015)

This data was used to conclude that the relative number of MCM complexes loaded during G1 regulates origin firing timing during S phase based on results that only used the relative number of MCMs at origins (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015).

This model was next used to illustrate that early-firing origins have multiple MCM complexes loaded. Different origins were engineered into the TALO8 plasmid system and a single binding site for the zinc-finger DNA binding protein Zif268 was introduced. MCM2 and ORC2 were tagged with the HA epitope and expressed an HA-tagged Zif268 (Das, Borrman, Liu,

Bechhoefer, & Rhind, 2015). The TALO8 plasmids containing different origins were purified and Western blotting was used to determine how many MCM complexes were loaded relative to the Zif268 control (Fig. 3A,B). After providing evidence that multiple MCMs can be loaded on a single origin and affect firing time, this model was used to show that reducing the number of MCM complexes loaded, delayed the firing time of the affected origin (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015).

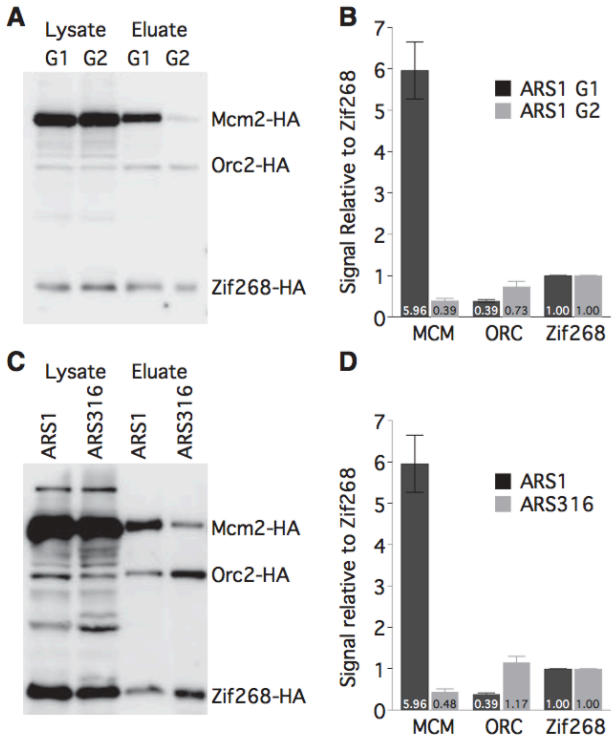

Figure 3. ARS1 has multiple MCM complexes loaded in vivo. Reprinted from Das et. al. (2015)

The evidence from this model presents a mechanistic overview of how replication is timed and regulated in *Saccharomyces cerevisiae* (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015). It supports the model of replication timing as a stochastic event with competition at origins for rate-limiting activators. The experimental data suggests that origins that compete more efficiently for limiting activators are more likely to fire early, and thus, replicate early. The number of MCM complexes loaded at origins contributes to this competition and leads to the efficiency and timing of origin firing during S phase. One MCM is no more likely to fire than another, but increasing the number of MCM complexes at an origin increases the likelihood

that it will fire earlier than an origin with fewer complexes. Therefore, this model demonstrates a “biochemically plausible mechanism for regulating origin efficiency and timing” based on experimentation with a large population of cells (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015).

Hypotheses of the Stochastic Model

Although the stochastic model illustrates how replication timing is regulated in a population array, it leads to important questions about how MCM complexes are loaded at single origins. As stated previously, there are two hypotheses for which MCM loading may be regulated: the rate at which MCMs are loaded or the capacity for MCM loading (Yang, Rhind, & Bechhoefer, 2010). If MCM loading is rate dependent due to the specific activity of ORC determining how many complexes are loaded, with no effect from crowding, then a Poisson distribution of MCM numbers is expected (Birnbaum, 1954). A skewed bell-shaped curve with a rightward tail would be expected (Figure 4A) (Das & Rhind, 2016). However, if the capacity for MCM complexes to be loaded is limited, with high capacity origins able to load more MCMs than low capacity origins, a saturation model would be expected. ORC would load as many MCMs as will fit at an origin, but then add additional MCMs more slowly (Figure 4B) (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015).

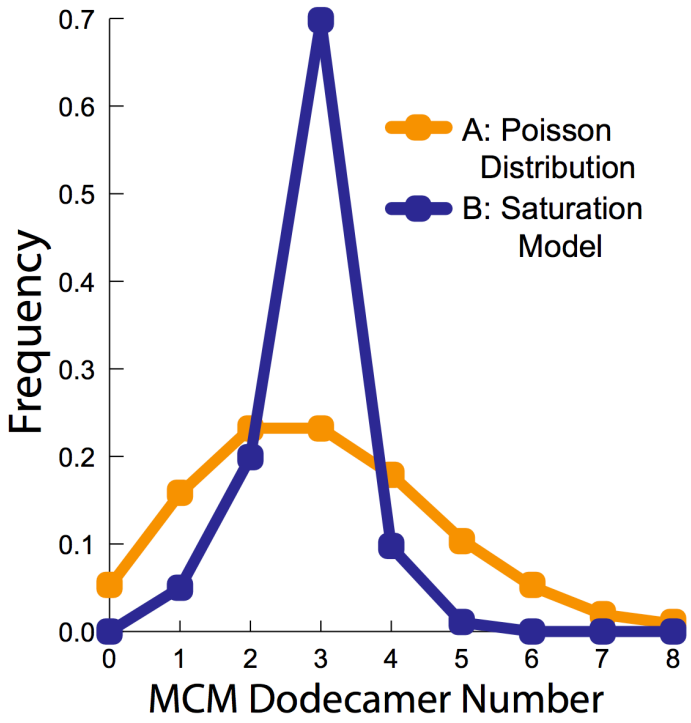

Figure 4. Possible MCM distributions at a known early origin- ARS1, with an average of about 3 MCMs.
A) Poisson distribution, expected if MCM loading based on rate of loading by ORC.
B) Saturation model, expected if MCM loading based on capacity for MCMs.

Single Molecule Biochemistry

Previously cited biochemical analysis of replication timing provided the multiple MCM model and the stochastic hypotheses. However, the data was limited in that it only estimated the average number of MCMs loaded at origin (Das, Borrman, Liu, Bechhoefer, & Rhind, 2015). Single molecule biochemistry can be implemented with fluorescent, single-molecule counting of MCMs loaded in vivo to measure the distribution of the number of MCMs loaded on individual origins (Friedman & Gelles, 2015).

To confirm the MCM complex model on a single molecule level, specific biochemical approaches can be implemented, including origin isolation by utilizing the interaction between the lac operator (lacO) and the lac repressor protein, LacI (Forde, Ghose, Slater, Hine, Darby, & Hitchcock, 2006). In its natural function, the lac repressor acts through a helix-turn-helix motif located in its DNA binding domain (Schumacher, Choi, Zalkin, & Brennan, 1994). The domain binds base-specifically to the major groove in the operator region with residues of related hinge alpha helices binding to base contacts in the minor groove (Schumacher, Choi, Zalkin, &

Brennan, 1994). This binding increases the affinity of RNA polymerase for the promoter region, disallowing for dissociation and preventing transcription of the mRNA encoding the Lac proteins (Daber, Stayrook, Rosenberg, & Lewis, 2007). This system may be amplified, with multiple copies of lacO stably integrated into a eukaryotic genome- developing a binding site array for LacI (Single Cell Manipulations).

The yeast strain containing the TALO8 plasmid with lacO and ARS1 is also transformed with MCM4-GFP- allowing for fluorescently labeled MCM complexes to be loaded to the single plasmid for study. Pull down of this TALO8 plasmid is achieved first by using a SNAP-tag. SNAP-tag is a 20 kDa mutant of the DNA repair protein O^6-alkylguanine-DNA alkyltransferase that covalently binds specifically and rapidly with benzylguanine (BG), labeling the SNAP-tag with a synthetic probe (Figure 5) (SNAP-tag Technologies: Novel Tools to Study Protein Function). This tagging system has a variety of advantages. First, the rate at which the SNAP-tag binds to BG is independent of the synthetic probe attached to BG (SNAP-tag Technologies: Novel Tools to Study Protein Function). Next, there are no restrictions on expression host with the SNAP-tag system. Finally, the SNAP-tag products are chemically inactive towards other proteins, eliminating nonspecific binding (SNAP-tag Technologies: Novel Tools to Study Protein Function).

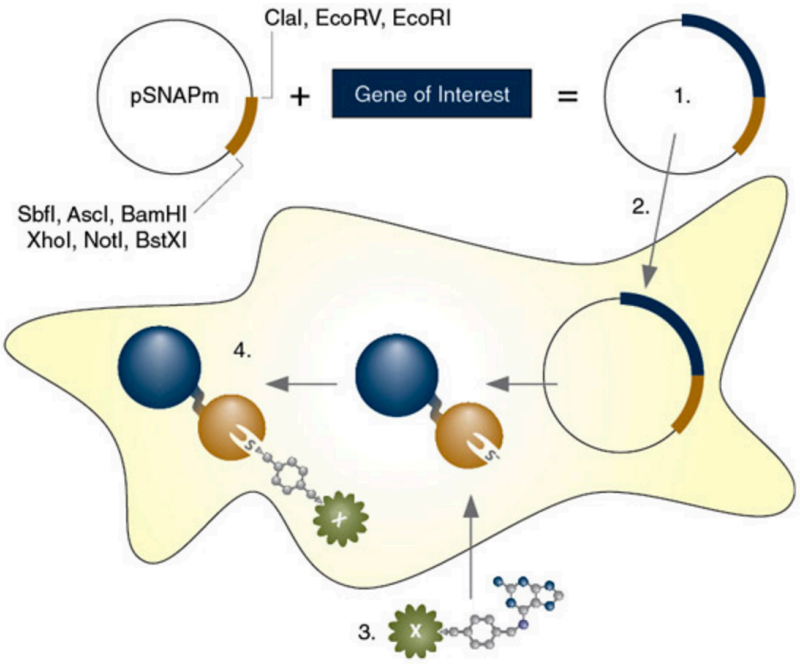

Figure 5. Imaging with SNAP-tag Technology: Clone gene of interest into NEB expression vector. 2) Transfect plasmid fusion into cells, protein is expressed in cells. 3) Add label of interest. 4) Covalent modification occurs, labeling protein for visualization. Reprinted from (SNAP-tag Technologies: Novel Tools to Study Protein Function)

The next feature of the pull down is the synthetic probe attached to BG. Along with a 649-fluorophore for labeling, biotin is attached to BG. When the flow cell is coated in streptavidin, the well-characterized relationship between biotin and streptavidin pulls the entire single molecule complex down. This interaction is one of the strongest, non-covalent interactions (Chivers, Koner, Lowe, & Howarth, 2011). The high binding affinity is due to several chemical interactions. First, there is a complementarity between the binding pocket of streptavidin and biotin with 8 hydrogen bonds made to residues within the binding site (DeChancie & Houk, 2007). Next, there is a 'second shell' involving hydrogen bonding to residues within the first shell and numerous van der Waals forces made to the biotin when in the streptavidin pocket (DeChancie & Houk, 2007). Finally, the affinity between streptavidin and biotin is influenced by "stabilization of a flexible loop connecting B strands 3 and 4 (L3/4), which closes over the bound biotin, acting like a 'lid' over the binding pocket and contributing to the extremely slow biotin dissociation rate" (DeChancie & Houk, 2007).

In implementing the aforementioned biochemistry, single plasmid molecule pull down can be achieved and used to count the number of MCMs loaded on a single origin, thus testing the hypothesis that origins that fire early in S phase have more MCMs loaded than origins that fire late in S phase. The complete outline of this process may be seen in Figure 6.

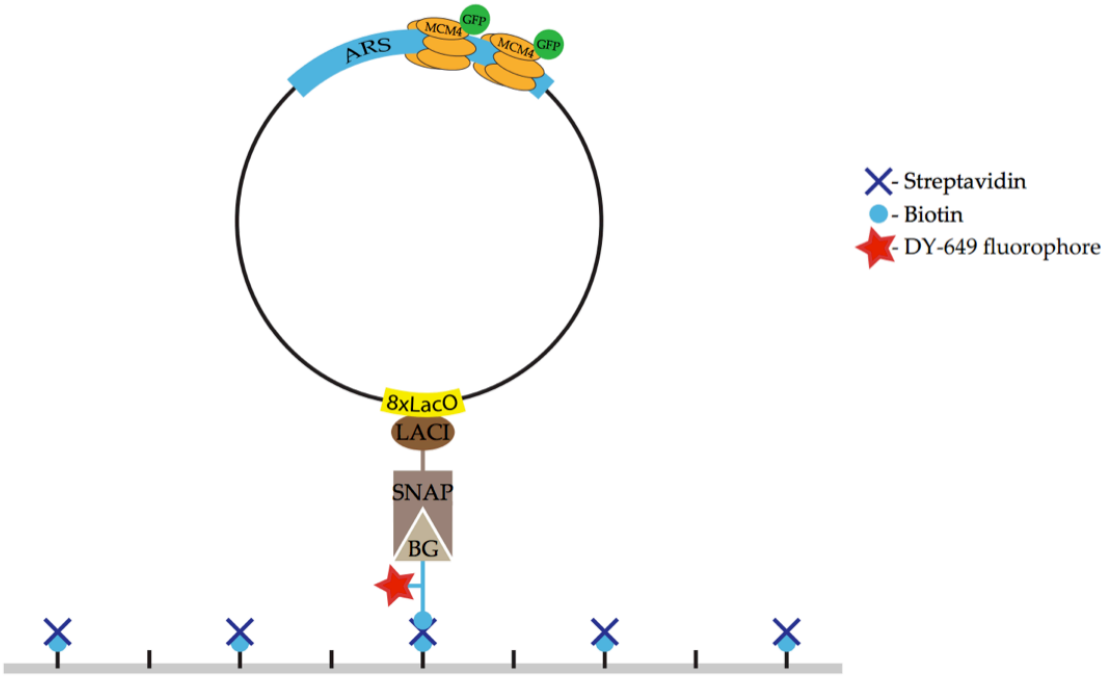

Figure 6. Single molecule pull down to quantify MCM complexes

MATERIALS

Table 1: Strains and plasmids used in this project

Strain Number	Description
yFS833	MCM4 Wild Type
yFS930	MCM4-GFP
yFS961	LacI-SNAP (cNAT)
yFS977	MCM4-GFP and LacI-SNAP
yFS979	MCM4-GFP, LacI-SNAP, and TALO8
yFS980	MCM4-mNeonGreen and LacI-SNAP
yFS981	MCM4-mNeonGreen, LacI-SNAP, and TALO8
yFS989	MCM4-GFP
yFS990	MCM4-GFP and LacI-SNAP
yFS991	LacI-SNAP
MMY1198	SEC3-SNAP
pFS270	GFP-HPH Cassette
pFS449	yeGFP-HPH Cassette
pFS454	mNeonGreen-KAN Cassette
pFS455	mNeonGreen-HPH Cassette
pFS458	LacI-FLAG Cassette
pFS466	SNAP Cassette

Table Two: Primers used in this project

Primer Name	Sequence (5'-3')	Use
MH2r	cgcacttaacttcgcatctgTTATTAATTGTTACGCAGGGAATGATTGTAGTAGACAGCA	Checking primer for GFP integration in MCM4
MH7r	CGAGGGTGTAAGGAGATCAGTTCGCCTGAATAACCGTGTCggtgacggtgctggtttaattaac	Checking primer for GFP integration in MCM4
KN07	AATCAGCTGTTGCCCGTCTC	Checking primer for integration of SNAP.NAT
KN13	TGGTGAAGGACCCATCCAGT	Checking primer for integration of SNAP.NAT downstream of LacI
MH08	TTATTAATTGTTACGCAGGGAATGATTGTAGTAGACAGCAtgggcagatgatgtcgagg	To isolate GFP-HPH construct
KN09	CGAGGGTGTAAGGAGATCAGTTCGCCTGAATAACCGTGTC aacagtaaaggagaagaact	To isolate GFP-HPH construct
LD200	GTCTTCTGATATCCAGGAAG	Checking primer inside MCM4 for GFP
LD202	CGTTGCCTCATCAATGCGAG	Checking primer inside ARS1 for pFS408
LD203	CAGTGAGCGCAACGCAATTA	Checking primer for pFS408
LD206	CTCGCATTGATGAGGCAACG	Checking primer for pFS408
LD207	AGTTCCTCGGTTTGCCAGTT	Checking primer for pFS408
MH4	CGGCACCGACTTTACCATAG	Checking primer in MCM4 for GFP
LD223	cggtaatacggttatccacag	Fwd primer pFS458.1
LD224	caccgcatagggtaataact	Rev primer pFS458.2
LD222	agttattaccctatgcggtg gacggtatcgataagcttga	Fwd primer pFS458.2
LD227	CATTTCGCAGTCTTTGTCCAT CTT TGG TGG AGT ACA GGA TCC	Rev primer pFS458.2
LD225	GGATCCTGTACTCCACCAAAG ATGGACAAAGACTGCGAAATG	Fwd primer pFS466
LD226	ctgtggataaccgtattaccg TTAACCCAGCCCAGGCTT	Rev primer pFS466
LD215	CGAGGGTGTAAGGAGATCAGTTCGCCTGAATAACCGTGTC GGT GCT GGA GCA GGT GCA GGA GCT GGT GCT aacagtaaaggagaagaact	Checking primer for MCM4-GFP 21bp linker

METHODS

Verification of Strains and Quality Control

First, YFS961, YFS930, and MM1198 single colonies were streak-plated on YPD media, and the knockout plates: –Trp, -Ura, YES, YES NAT, -His, and –Ade. Also, using YFS961 and YFS930, liquid cultures were made by transferring a single colony into 5mL of YPD and spinning overnight at room temperature. The knockout plates were observed and the overnight cultures were used to check for fluorescence via microscopy.

Genomic DNA Preparation

YFS961 and YFS930 underwent genomic preparation in order to serve as templates for PCR amplification. For this process, these candidates were once again cultured overnight in liquid YPD. The next day, 1.5ml of each was pelleted by centrifugation at 20,000 x g for five minutes. The pellets were resuspended in 200ul of cell lysis buffer and immersed in a dry ice-ethanol bath for two minutes. Both tubes were then transferred to a 95°C water bath for one minute. This process was repeated and then 200uL of chloroform was added. The microcentrifuge tubes were centrifuged at 20,000 x g and the supernadent was removed. Several ethanol washes were then preformed and the pellet was dried at room temperature for five minutes. Finally, the genomic DNA was resuspended in 40ul of water. A cleanup was performed on the genomic DNA with binding and wash buffer in a column purification that yielded 20ul. This DNA was then used in Taq PCR reactions with the primers MH2r and MH7r for YFS930 and KN7 and KN13 used with YFS961. Finally, a gel electrophoresis with 1% agarose gel was performed in order to verify the candidates. A freezer stock of YFS930 was obtained and the genomic DNA preparation and PCR was repeated in order to move forward with this candidate.

Plasmid MiniPrep

Next, the GFP-cassette plasmid, PFS270, was obtained from a freezer stock, streaked on an LB Carb plate, and used to create an overnight culture. 1.5mL of the bacteria was pelleted at 6,000 x g for one minute at room temperature then resuspended in 200ul of Solution A (50mM

Tris pH 7-8, 10mM EDTA pH 8.0) and had 5ul of 10mg/ml RNase A added. After vortexing, 200ul of Solution B (200mM NaOH, 1% SDS) was added and the mixture inverted. 300ul of Solution C (3M KOAc pH 5) was added and then the mixture was pelleted by centrifugation at 16,000 x g for five minutes. Once done, 450ul of the supernatant was transferred to a new tube, and the DNA was ethanol precipitated by adding 2.5V (1,125ul) of 100% EtOH. After centrifugation, another ethanol wash was preformed using 70% EtOH and then the pellet was dried and resuspended in 50ul water. Finally, the plasmid DNA was cleaned up using column purification and eluting with 20ul of water.

Yeast Transformation

Using PFS270, the GFP-HPH cassette was amplified using PCR techniques with the primers MH08 and KN09. 400ul of the PCR product was cleaned up using column purification and a 50ml overnight culture of YFS961 was inoculated. The overnight culture was then spun down at 3,000rpm for three minutes at room temperature. The pellet was resuspended in 5ml of sterile TE buffer and spun down. It was then resuspended in 5ml of LiAc Mix and spun down. Then, the pellet was once more resuspended in 250ul of LiAc Mix and 100ul was aliquoted into an epindorph tube for one transformation. Next, 10ul of the clean PFS270 PCR product was added to the cells along with 10ul of 10mg/ml Salmon Sperm DNA. 700ul of PEG Mix was added and the tube was vortexed and then incubated for 30 minutes at 30°C. A heat shock was then preformed for 15 minutes at 42°C and then spun down, removing the supernadent. The cells were resuspended in 300ul of TE and then plated on YPD. The next day the cells were replica plated onto YPD HPH.

Bacteria Transformation

A bacterial transformation was preformed in order to obtain a different fluorescent plasmid candidate, PFS449, that contains a GFP derivative. In order to do this, a tube of dh5-alpha E. coli competent cells was thawed on ice for 10 minutes. 50ul of the cells were then pipetted into a microcentrifuge tube and 1ul of diluted plasmid 449 DNA was added. This mixture was then placed on ice for 30 minutes and then heat shocked for 30 seconds at 42°C. It

was placed back on ice for five minutes and then 950ul of room temperature LB media was added. The tube was then spun for one hour at 37°C. Several 10-fold serial dilutions of this mixture were then plated on LB Carb and grown overnight at 37°C. The next day, overnight cultures of six candidate colonies were made and a miniprep was preformed in order to obtain the plasmid DNA. 1ul of the DNA was run on a 1% agarose gel in order to determine which of the six candidates contained plasmid DNA. The candidates that contained DNA were then digested with several restriction enzyme: SAP1 in order to linearize, HPA1 to create a double stranded cut, and BST11 to create a different double stranded cut. Finally, a 1% agarose gel of the digested products was run in order to determine which of the candidates truly contained PFS449.

Noodle-Making

Yeast 'noodles' of the experimental strain were created in order to be ground down and used as yeast extract. To start, 2L of media was inoculated with the appropriate strain from a starter culture and grown until an OD600 of 1.2-1.5. The cells were spun down in 500ml containers at 3000 rpm for 10 minutes at 4°C. The pellet was resuspended in 25mls ice-cold dH_2O and all resuspended yeast was combined into one 50ml conical tube. The cells were then spun down again at 3000 rpm for 5 minutes at 4°C. The supernadent was discarded and the cells were washed with 50mls of ice-cold dH_2O, then they were spun at 3000 rpm for 5 minutes at 4°C. Next, the yeast paste was transferred to a 5ml syringe and expunged into a new 50ml conical tube containing 25ml of liquid N_2. The excess liquid N_2 was removed and the yeast noodles were stored at -80°C until use.

Ball Mill Grinding of Yeast Noodles

Using the Retsch Ball Mill Grinder, previously made yeast noodles were ground into a yeast extract for use in experimentation. First, the ball mill contained was cooled down by pouring liquid N_2 over it until a 'bubbling' effect appeared. The yeast noodles were put into the container and the cooling process was repeated. The apparatus was placed into the Retsch machine, locked into place, and ground for 1m30s at 400 rpm. The container was placed back

into liquid N_2, the yeast was dislodged from the sides, and the cooling process was repeated. These steps were repeated until the sample was ground a minimum of 6 times and had a powdery appearance.

These protocols were repeated for strains yFS833 and yFS989 and for plasmids pFS458 and pFS466.

RESULTS

Strain Engineering

At the beginning of this project the strain yFS961 was originally used as the LacI-SNAP strain that MCM4-GFP was transformed into from pFS270. At the same time, MCM4 was tagged with mNeonGreen from both pFS454 and pFS455 and transformed into yFS961 in order to compare the strength of fluorescence between the two fluorophores- GFP and mNeonGreen. The MCM4-GFP and LacI-SNAP strain became yFS977 and the MCM4-mNeonGreen from pFS454 and LacI-SNAP strain became yFS980. Next, the TALO8 plasmid was transformed into yFS977 and the strain was labeled yFS979. Finally, TALO8 was transformed into yFS980 and the new strain was named yFS981.

Next, experimentation moved forward with yFS979. Yeast noodles were made, the ball-mill grinder was used to make yeast extract, the BG-biotin-649 biotin-fluor was conjugated, and the complex was run over flow-cells functionalized with biotin and streptavidin.

At this point in the project we began the process over due to the presence of a previously annotated LacI-FLAG present in yFS961 that would compete with LacI-SNAP and complicate the experimental setup. Starting over began by fluorescently labeling MCM4 by PCR-based cassette tagging with GFP from pFS270 using the primers MH08 and LD215. The plasmid piece was then transformed into yFS833 that contained MCM4, this strain then became known as yFS989 (Figure 6).

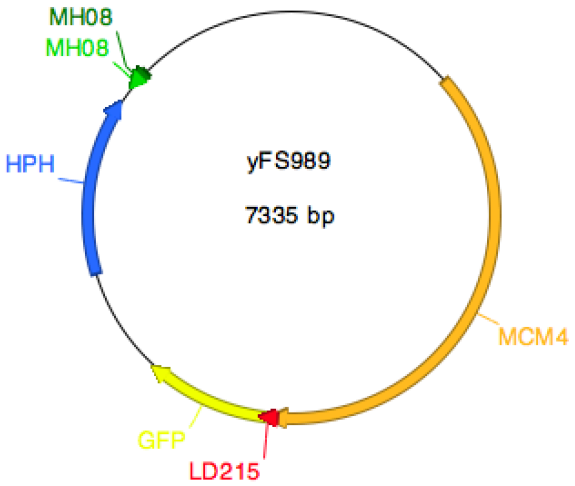

Figure 7. yFS989 strain with MCM4-GFP

To create the LacI-SNAP strain, PCR-based cassette tagging was used to tag a CMV promoter and LacI from the plasmid pFS458 using the primers LD223 and LD224. Using the primers LD222 and LD227, the pUC origin, ampicillin resistance marker, and the functional URA3 gene were also PCR'd from pFS458. The SNAP-tag was PCR'd from the plasmid pFS466 using the primers LD225 and LD226. A Gibson Assembly combined the three PCR products to create the LacI-SNAP plasmid that was then transformed into E. coli. The plasmid construction process and confirmation with restriction enzyme EcoRV can be seen in Figure 7. A summary of the transformation process is illustrated in the flow chart in Figure 8.

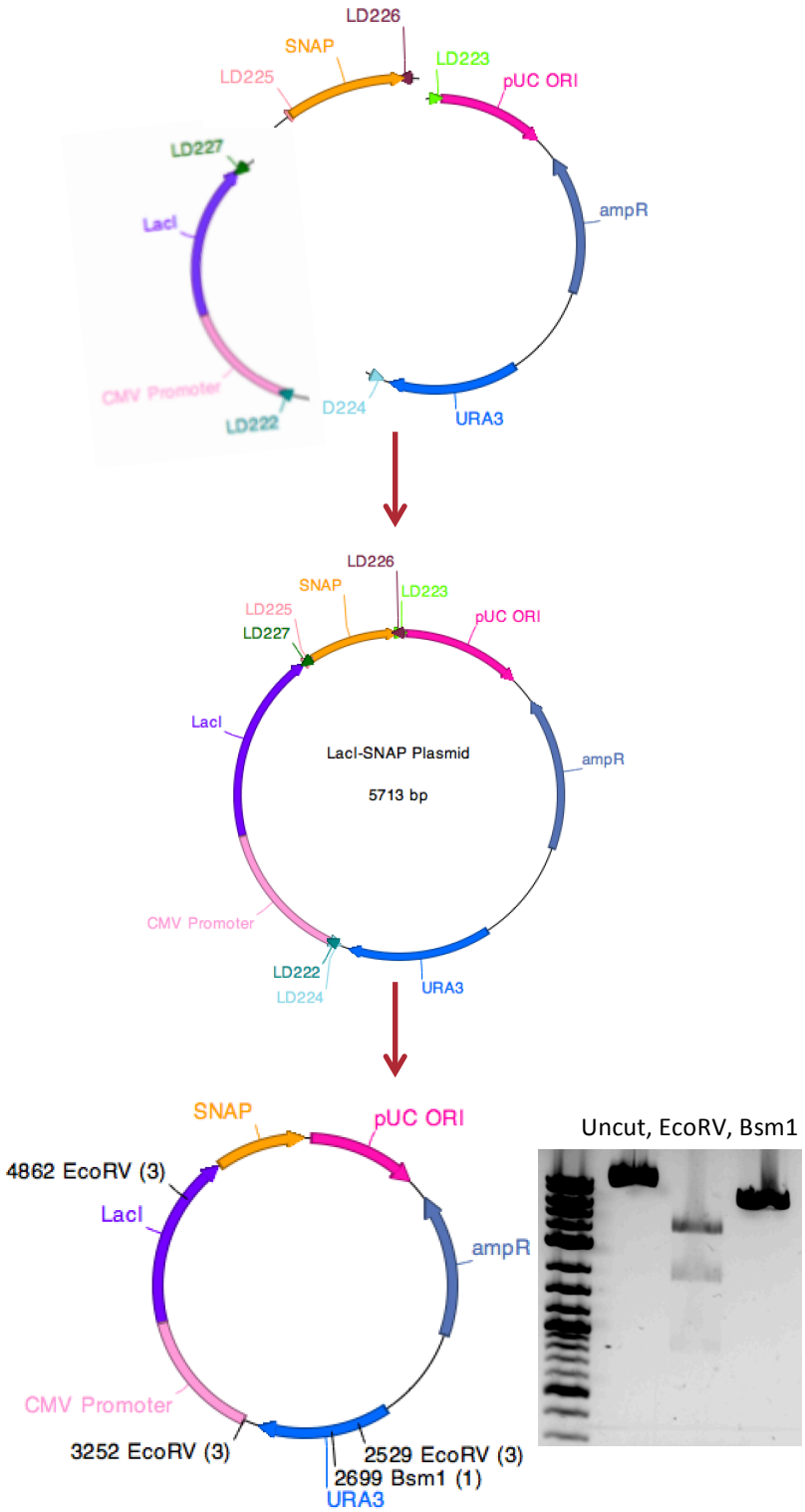

Figure 8. Plasmid construction process and confirmation

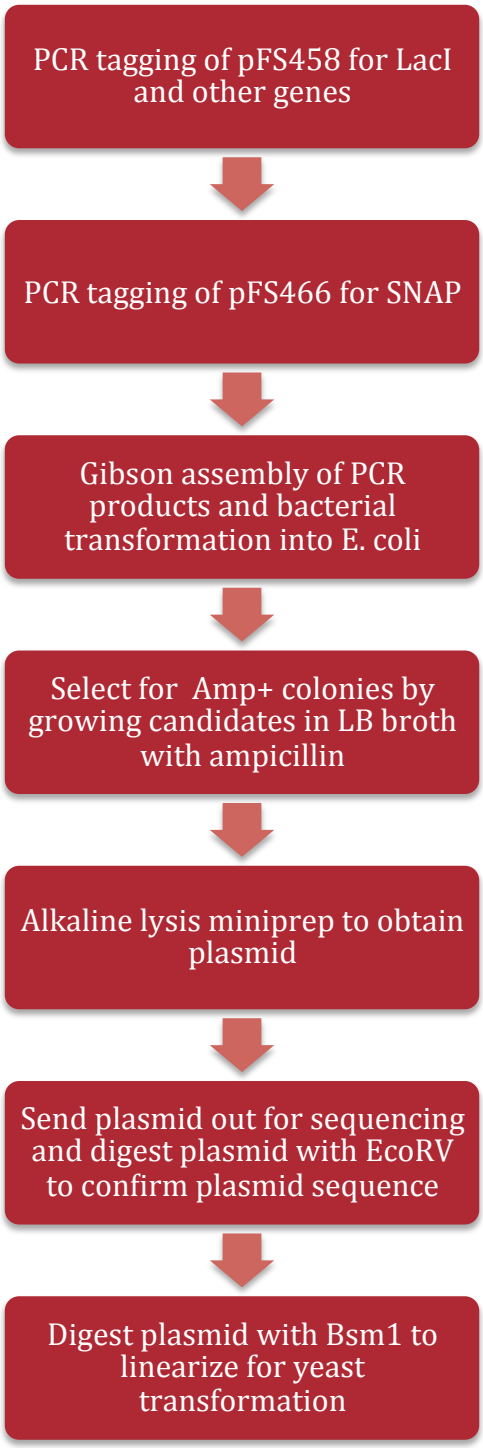

Figure 9. Summary of transformation

The LacI-SNAP plasmid was then transformed into yPF989 to create the strain yFS990.

DISCUSSION

The premise of this project was to be able to quantify the number of MCM complexes loaded on plasmids containing single, well-characterized replication origins from G1 arrested yeast. This goal was to be accomplished by first producing a yeast strain containing SNAP-tagged LacI, and MCM4 tagged with GFP. Then, a TALO8 plasmid would be inserted that contained the well-characterized origin, ARS1, eight copies of the lacO high affinity binding site, and a functional tryptophan gene for selective propagation. The lacO binding site for the LacI repressor would allow for affinity purification of the plasmids with SNAP-tagged LacI because the SNAP tag allows for covalent attachment of a biotin-fluor tag called BG-biotin-649. In conjugating this tag to yeast extracts, LacI would be functionalized with both biotin for attachment to a flow-cell surface and a red fluor for visualization. This system would purify single molecules and tether them to a flow-cell; GFP labeled MCMs could then be counted by way of quantitative photobleaching using TIRF microscopy.

Originally, the project was based off of the strain built from yFS961, but the purification system failed and MCMs could not be quantified for several reasons. Not only was the purification system not yet optimized for the complex, but also unreacted dye could not be removed from the yeast extract and purification was complicated.

It was at this point in the process that yFS833 was used in conjunction with pFS270 to make yFS989 and pFS458 and pFS466 were used to create the LacI-SNAP plasmid that we were able to move forward with. In changing the MCM4-GFP tagging, we inserted a 10 amino acid linker between the genes in order to increase the functionality of the GFP tag. In creating a new LacI-SNAP plasmid, we were able to confirm that these genes were within reading frame, allowing for the LacI-SNAP fusion protein to be translated correctly once inserted into the yeast genome.

For the future directions of this project, the TALO8 plasmid will be transformed into yFS990 and this strain will be used to optimize extract purification. Once the purification is successful and single plasmid molecules can be immobilized, TIRF microscopy will be used to quantify the number of GFP fluorescent MCM4s bound to ARS1 by way of photobleaching. With single-molecule quantitation optimized, not only will new yeast strains will be produced that

contain mutations to the B1 and B2 elements within known origin sequences, but also known later-firing ARS origins will be implemented in the system in order to determine the regulation of MCM loading.

Determining the number of MCMs bound to specific origins, both endogenous and mutated, will help test the hypothesis that the number of MCMs loaded on origins of replication determines firing efficiency and timing. This process is significant because the timing of genome replication in S phase has been correlated with gene expression, cellular differentiation, development, and DNA repair. Orderly replication is necessary to not only maintain genome stability, but also regulate the events of genome metabolism to conduct the chemical processes that maintain all living cells.

BIBLIOGRAPHY

Bechhoefer, J., & Rhind, N. (2012). Replication timing and its emergence from stochastic processes. *PMC , 28* (8), 374-381.

Bell, S. P., & Kaguni, J. M. (2013). Helicase loading at chromosomal origins of replication. *Cold Spring Habor Perspectives in Biology , 5* (6).

Birnbaum, A. (1954). Statistical methods for Poisson processes of exponential populations. *Journal of the American Statistical Association , 49*, 254-266.

Brewster, A. S., Wang, G., Yu, X., Greenleaf, W. B., Carazo, J., Tjajadi, M., et al. (2008). Crystal structure of a near-full-length archaeal MCM: Functional insights for an AAA+ hexameric helicase. *PNAS , 105* (51), 20191-20196.

Chivers, C., Koner, A., Lowe, E., & Howarth, M. (2011). How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimeric tetramer. *Biochemical Journal , 435* (1), 55-63.

Czajkowsky, D. M., Liu, J., Hamlin, J. L., & Shao, Z. (2008). DNA combining reveals intrinsic temporal disorder in the replication of yeast Chromosome VI. *Journal of Molecular Biology , 375* (1).

Daber, R., Stayrook, S., Rosenberg, A., & Lewis, M. (2007). Structural analysis of Lac repressor bound to allosteric effectors. *Journal of Molecular Biology , 370*, 609-17.

Das, S. P., Borrman, T., Liu, V., Bechhoefer, J., & Rhind, N. (2015). Replication timing is regulated by the number of MCMs loaded at origins. *Genome Research , 25* (12).

Das, S., & Rhind, N. (2016). How and why multiple MCMs are loaded at origins of DNA replication. *Bioessays , 38*, 613-617.

DeChancie, J., & Houk, K. (2007). The origins of femtomolar protein-ligand binding: hydrogen bond cooperativity and desolvation energetics in the biotin-streptavidin binding site. *Journal of the American Chemical Society , 129* (17), 5419-29.

Forde, G. M., Ghose, S., Slater, N., Hine, A., Darby, R., & Hitchcock, A. (2006). LacO-LacI interaction in affinity adsorption of plasmid DNA. *Biotechnology and Bioengineering , 95* (1).

Friedman, L. J., & Gelles, J. (2015). Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. *Methods , 86*, 27-36.

Goren, A., & Cedar, H. (2003). Replicating by the clock. *Nature Reviews Molecular Cell Biology , 4* (1).

Hiraga, S.-I., Alvino, G., Chang, F., Lian, H.-Y., Sridhar, A., Kubota, T., et al. (2014). Rif1 controls DNA replication by directing protein phosphatase 1 . *Genes and Development , 28* (4).

Kalejta, R., & Hamlin, J. (1996). Composite patterns neutral/neutral two-dimensional gels demonstrate inefficient replication origin usage. *Molecular and Cell Biology , 16* (9), 4915-4922.

Patel, P. K., Arcangioli, B., Baker, S. P., Bensimon, A., & Rhind, N. (2006). DNA repliation origins fire stochastically in fission yeast. *Molecular Biology of the Cell , 17* (1).

Rhind, N., & Gilbert, D. M. (2013). DNA replication timing. *Cold Spring Habor Perspectives in Biology , 5* (8).

Rhind, N., Yang, S., & Bechhoefer, J. (2010). Reconciling stochastic origin firing with defined replication timing. *Chromosome Research , 18* (1).

Sakakibara, N., Kelman, L. M., & Kelman, Z. (2009). Unwinding the structure and function of the archaeal MCM helicase. *Molecular Microbiology , 72* (2), 286-296.

Schumacher, M., Choi, K., Zalkin, H., & Brennan, R. (1994). Crystal structure of LacI . *Science , 266* (5185), 763-70.

Single Cell Manipulations. (n.d.). (DKFZ) Retrieved April 1, 2017, from BioQuant: https://malone.bioquant.uni-heidelberg.de/methods/single_cell/single_cell.html

SNAP-tag Technologies: Novel Tools to Study Protein Function. (n.d.). Retrieved April 1, 2017, from NEB: https://www.neb.com/tools-and-resources/feature-articles/snap-tag-technologies-novel-tools-to-study-protein-function

UniProtKB - P32561 (RPD3_YEAST). (n.d.). Retrieved April 1, 2017, from UniProt: http://www.uniprot.org/uniprot/P32561

Yang, S., Rhind, N., & Bechhoefer, J. (2010). Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Molecular Systems Biology , 6* (1).

Yeeles, J., Deegan, T., Janska, A., Early, A., & Diffley, J. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature , 519*, 431-435.